

READ ME

The Excel Data file `kraft2023effectsizes.xlsx`. contains the data used in the paper:

Kraft, M.A. (Forthcoming). The effect size benchmark that matters most: Education interventions often fail. *Educational Researcher*. <https://doi.org/10.3102/0013189X231155154>

Unfortunately, these data do not contain standard errors because most of the sources I used to compile the effect sizes did not report this information. If you need standard errors, I recommend using the What Works Clearinghouse databased described below.

There are 29 effect sizes without citations. These come from a summary report of the Investing in Innovation Evaluations (i3) grants. These were effect sizes report to IES but never published.

Boulay, B., Goodson, B., Olsen, R., McCormick, R., Darrow, C., Frye, M., ... & Sarna, M. (2018). The Investing in Innovation Fund: Summary of 67 Evaluations. Final Report. NCEE 2018-4013. *National Center for Education Evaluation and Regional Assistance*.

As I describe below in the online data appendix, it is possible the data contain duplicates or errors given that we drew from sources the surveyed overlapping literatures but used different reporting norms.

I would be grateful if you would share any errors you may find in the data. You can reach me at mkraft@brown.edu.

Online Data Appendix

Data Sources

I use six main sources for this expanded database of effect sizes from randomized control trials of education interventions with standardized achievement outcomes. These sources consist of third-party collections of effect-size estimates (4 U.S. based; 1 U.K. based) and a handbook chapter by Roland Fryer reviewing randomized field experiments of human capital interventions. The vast majority of new effect sizes I add are from one source: the Institute of Education Sciences (IES) What Works Clearinghouse database. In addition, I also found additional studies reviewed by The Education Endowment Foundation and The Best Evidence Encyclopedia since I last accessed these sources in 2018. I provide detailed descriptions of each data source with further summary statistics from this expanded dataset in Table A1.

Data Construction

Dr. Simpson was kind enough to provide me with an advanced copy of his technical comment, which pointed out several small data entry errors. This prompted my research team and I to reconstruct our entire expanded analytic dataset to ensure it unambiguously meet my three inclusion criteria: (1) education interventions (2) from randomized controlled trials (3) with a standardized test outcome.

In this process, we worked to remove any effect sizes that employed an ambiguous experimental research design or appeared as duplicates in the data. We found that some effect sizes from the Best Evidence Encyclopedia evidence reviews used ambiguous research design categories and may have included supposed experimental designs of varying causal rigor. For example, some of the studies compare treated groups to business-as-usual control groups but make no mention of randomization. We also found that the six main sources of effect sizes have a non-negligible degree of overlap. Citation and effect-size reporting norms are not consistent across, or even within, each source. We created a unique effect-size ID based on the last name of up to six authors, publication year, subject, grade level, effect size rounded to two digits, and outcome type to remove duplicates introduced as a result of one study being reported in multiple sources. We then hand reviewed the full dataset for possible duplicates.

The different reporting norms across the third-party sources make identifying and eliminating all duplicates a particular challenge. We believe our review process has minimized these duplicates but it is possible some duplicates still exist within the data. We identify a unique study ID based on the last name of up to six authors and publication year. To the extent that there are multiple studies published by the same author teams in the same year our approach will understate the total number of studies included. We also note that some research results reported in the Investing in Innovation Evaluations report did not include researcher identification information.

Codes

After recompiling this expanded analytic sample, my research team and I coded these data for a range of characteristics including study sample size, grade level, subject, and whether a test was narrow or broad. I created indicator variables to identify which grades each study focused on. Many of the interventions ranged across multiple grades and only presented overall effect sizes. In these cases, I include effects sizes in all grade-level groups that are represented in each sample. In cases where effect sizes were listed separately by grade, I included them as separate observations. The result is that effect sizes are not mutually exclusive by grade across the sample.

Following Hill et al. (2007), I distinguish between standardized tests that cover a broad subject matter and narrow standardized tests that are either the constituent parts of the broad composite score or specific subject tests. Studies often report effect sizes for broad and narrow standardized outcomes for the same sample. I include only broad standardized measures when both broad and narrow effect sizes are reported to ensure these non-independent effect sizes are not double-counted. It is possible that different reporting norms across sources caused both narrow and broad effect sizes from a single study to be included.

Table A1. *Description of sources used to collect effect size outcomes*

Source	Description	Mean	Mean (weighted)	Percentiles			Effect Sizes	Studies
				30th	50th	70th		
Fyer (2017) in the Handbook of Field Experiments, Vol. 2	Handbook of Field Experiments, Volume Two explains how to conduct experimental research, presents a catalog of research to date, and describes which areas remain to be explored. Chapter two looks at the findings from 196 randomized field experiments specifically in education. Citation: Fryer Jr, R. G. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In Handbook of economic field experiments (Vol. 2, pp. 95-322). North-Holland.	0.16	0.03	0.02	0.09	0.19	314	182
Best Evidence Encyclopedia	The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education under funding from the Institute of Education Sciences, U.S. Department of Education. It is intended to give educators and researchers fair and useful information about the strength of the evidence supporting a variety of programs available for students in grades K-12. I rely on the full evidence reports which are available for math, reading, science and early childhood can be downloaded at http://www.bestevidence.org/index.cfm	0.12	0.07	0.02	0.10	0.17	954	318

IES WWC Database	<p>The What Works Clearinghouse is an investment of the Institute of Education Sciences (IES) within the U.S. Department of Education that was established in 2002. The work of the WWC is managed by a team of staff at IES and conducted under a set of contracts held by several leading firms with expertise in education, research methodology, and the dissemination of education research. The WWC makes findings from reviewed studies available here https://ies.ed.gov/ncee/wwc/StudyFindings</p>	0.26	0.09	0.05	0.1 4	0.3 2	1477	270
IES Commissioned RCTs 2002-2013	<p>This report published by the Coalition for Evidence-Based Policy highlights key findings from 90 interventions that have been evaluated in IES-commissioned RCTs. The report can be found here http://coalition4evidence.org/wp-content/uploads/2013/06/IES-Commissioned-RCTs-positive-vs-weak-or-null-findings-7-2013.pdf</p>	0.02	0.01	-0.05	0.0 3	0.0 7	94	29
Investing in Innovation Evaluations	<p>Evaluations from the Investing in Innovation Fund, which provides competitive grants to local education agencies and non-profits to implement and evaluate educational interventions. All interventions are evaluated by outside organizations. The detailed report can be found here https://ies.ed.gov/ncee/pubs/20184013/pdf/20184013.pdf</p>	0.06	0.04	-0.02	0.0 3	0.1 0	100	27

Education Endowment Foundation	<p>The Education Endowment Foundation was established in 2011 by The Sutton Trust, as a lead charity in partnership with Impetus Trust (now part of Impetus - The Private Equity Foundation) with a £125m founding grant from the Department for Education. The EEF and Sutton Trust are, together, the UK government-designated What Works Centre for Education. The EEF conducts its own evidence reviews of the existing literature, meta- & meta-meta- analyses, and fund their own educational intervention program evaluations. Evidence summaries and program evaluations can be found here https://educationendowmentfoundation.org.uk</p>	0.15	0.02	0.01	0.0 8	0.1 7	487	236
--------------------------------------	---	------	------	------	----------	----------	-----	-----

Notes: Every effort was taken to remove duplicate studies and effect sizes across different data sources. Some duplicates may remain given coding difference across source types.